

Sequence analysis

A new *cis*-acting regulatory element driving gene expression in the zebrafish pineal gland

Shahar Alon¹, Eli Eisenberg^{2,*}, Jasmine Jacob-Hirsch³, Gideon Rechavi³, Gad Vatine¹, Reiko Toyama⁴, Steven L. Coon⁵, David C. Klein⁵ and Yoav Gothilf^{1,*}

¹Department of Neurobiology, The George S. Wise Faculty of Life Sciences, ²Raymond and Beverly Sackler School of Physics and Astronomy, ³Cancer Research Center, Sheba Medical Center, Tel Hashomer and Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel, ⁴Laboratory of Molecular Genetics and ⁵Program in Developmental Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA

Received on November 1, 2008; revised on December 21, 2008; accepted on January 8, 2009

Advance Access publication January 15, 2009

Associate Editor: Alfonso Valencia

ABSTRACT

Motivation: The identification of functional *cis*-acting DNA regulatory elements is a crucial step towards understanding gene regulation. *Ab initio* motif detection algorithms have been extensively used in search of regulatory elements. Yet, their success in providing experimentally validated regulatory elements in vertebrates has been limited.

Results: Here we report *in silico* identification and *in vivo* validation of regulatory elements that determine enhanced gene expression in the pineal gland of zebrafish. Microarray data enabled detection of genes that exhibit high expression in the pineal gland. The promoter regions of these genes were computationally analyzed in order to identify overrepresented motifs. The highest ranking motif identified is a CRX/OTX binding site, known to govern expression in the pineal gland and retina. The second highest ranking motif was not reported before; we experimentally validated its function *in vivo* by mutational analysis. The methodology presented here may be applicable as a general scheme for finding regulatory elements that contribute to tissue-specific gene expression.

Contacts: yoavg@tauex.tau.ac.il; elieis@post.tau.ac.il

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Identifying DNA regulatory elements in vertebrates is an important task for computational as well as experimental biologists in the post-genome era. Numerous computational methods, including *ab initio* motif detection programs, have been developed for the identification of putative regulatory elements (GuhaThakurta, 2006). Many of these methods rely on co-regulation of genes at the transcriptional level; promoter regions of co-regulated genes serve as an input and overrepresented strings among these regions are considered putative regulatory elements. However, there are only a few examples of experimental validation of computationally identified motifs in vertebrates (Blanchette *et al.*, 2006; Davies *et al.*, 2007; Wang *et al.*, 2006).

In the last decade, it has become possible to measure mRNA levels in a genome-wide scale using DNA oligonucleotide microarrays. This methodology provides a robust means for the detection of co-expressed genes (i.e. potentially co-regulated), creating another source of data for computational identification of putative regulatory elements (GuhaThakurta *et al.*, 2002; Roth *et al.*, 1998).

The key function of the pineal gland in all vertebrates is the nocturnal production and secretion of the hormone melatonin which has an important role in the regulation of many biochemical and physiological circadian rhythms. In non-mammalian vertebrates, the pineal gland contains a circadian clock that drives the circadian melatonin rhythm and is also a photoreceptive organ; photic signals directly synchronize the melatonin rhythm (Falcón, 1999). Proper development and function of the pineal gland depend on the expression of a specific set of genes that are involved in photoreception, phototransduction, melatonin synthesis and clock function. Revealing the regulatory mechanisms responsible for the expression of genes in a tissue-specific manner is an important step towards a comprehensive understanding of the developmental processes specifying the pineal identity. Three DNA binding sites that drive enhanced gene expression in the pineal are currently known: (i) The sequence TAATC/T, known as photoreceptor conserved element (PCE), which was found in the promoters of several pineal-specific genes (Appelbaum and Gothilf, 2006; Li *et al.*, 1998). This element serves as a binding site for the OTX5/CRX family of transcription factors which are important for pineal development (Appelbaum *et al.*, 2005; Chen *et al.*, 2007; Gamse *et al.*, 2002). (ii) The sequence TGACCCCAATCT, called pineal expression-promoting element (PIPE), which was found in the zebrafish *exo-rhodopsin* promoter (Asaoka *et al.*, 2002). (iii) The sequence combination GANNCTTA and TAAA, called pineal expression related element (PERE), recently found in the zebrafish *Rev-erbα* promoter (Nishio *et al.*, 2008). It is likely, however, that there are more regulatory elements that contribute to the expression of genes in a pineal-specific manner.

Here, transcripts that are highly expressed in the zebrafish pineal gland, including known and novel genes, were identified by comparing the transcriptome of the pineal gland with that of

*To whom correspondence should be addressed.

other tissues. Following an experimental validation of this list of transcripts, a computational approach was taken to identify regulatory elements that are important for the expression of genes in the pineal gland. In addition to identifying the PCE discussed above, this approach also identified a novel regulatory motif. *In vivo* mutation analysis confirmed that this motif directs expression of genes in the pineal gland, thereby proving the usefulness of the strategy taken here in detecting previously unrecognized regulatory elements.

2 METHODS

Pineal and brain RNA extraction: adult *Tg(aanat2:EGFP)Y8* transgenic zebrafish (Gothilf et al., 2002) which express enhanced green fluorescent protein (EGFP) in the pineal gland under the control of the *aanat2* regulatory regions, were used. Fish were anesthetized in 1.5 mM Tricane (Sigma, St. Louis, MO), sacrificed by decapitation, and pineal glands were removed under a fluorescent dissecting microscope. Since the pineal gland is a clock-containing organ and levels of certain transcripts may vary throughout the circadian cycle, glands were collected throughout the 24 h cycle at 4 h intervals. During the 24 h cycle, the fish were either maintained in 12 h light/12 h dark cycle (LD) or kept in constant darkness (DD). Twelve pineal glands were collected and pooled at each time point at each light condition and total RNA was extracted (RNeasy, Qiagen, Valencia, CA). In addition, total RNA was prepared from pools of brains, after the removal of the pineal gland and eyes. In order to account for changes in gene expression during the 24 h cycle pools of brains were either collected at ~12 a.m. or at ~12 p.m.

Preparation of labeled RNA and hybridization of microarrays: labeled RNA preparation and hybridization to microarrays was performed according to the Affymetrix manual with two-cycle target labeling protocol (http://www.affymetrix.com/support/downloads/manuals/expression_analysis_technical_manual.pdf). A total number of 14 Affymetrix microarrays were hybridized with RNA-pools of pineal glands from seven time points throughout the 24 h cycle for each light condition (LD or DD). Twelve Affymetrix microarrays were hybridized with brain RNA—six microarrays for each time point (~12 a.m. and ~12 p.m.).

Microarray data collected from GEO datasets and data submission: the following datasets were downloaded from GEO <http://www.ncbi.nlm.nih.gov/geo/>: three zebrafish muscle samples: GSM102669, GSM102671, GSM102674; five zebrafish heart samples: GSM112798, GSM112800, GSM112802, GSM112803, GSM112806; three zebrafish embryos samples GSM95623, GSM95624, GSM95625. The pineal and brain microarray data was submitted to the GEO: GSM333303–GSM333316 (pineal); GSM337618–GSM337620, GSM337624–GSM337626, GSM337631–GSM337633 and GSM337638–GSM337640 (brain).

Identifying promoters: only transcripts that are part of known genes were used. In the case that a promoter was not reported in the literature, the transcription start site (TSS) of the gene was estimated by aligning ESTs to the gene sequences using the UCSC genome browser (<http://genome.ucsc.edu/>). The 5' end of the resulting cluster was considered putative TSS and 500 bp upstream of the TSS and the 5'-untranslated region (UTR) were used as putative promoter.

Randomly chosen zebrafish promoters: random zebrafish promoters were chosen from all the zebrafish genes in the NCBI mRNA reference sequences collection (RefSeq). The promoter data was downloaded from the UCSC table browser (<http://genome.ucsc.edu/>).

***In vivo* transient expression:** the *Aanat2-EGFP-PRDM* construct is described elsewhere (Appelbaum et al., 2004). When injected into zebrafish embryos, this construct drives pineal-specific expression of EGFP. A two-nucleotide mutation (CAATC to CGGTC) was introduced into *Aanat2-EGFP-PRDM* using the QuickChange site-directed mutagenesis kit (Stratagene, La Jolla, CA) as instructed by the manufacturer and confirmed by DNA sequencing. *In-vivo* transient expression assays of the EGFP-containing

constructs were performed by microinjection of zebrafish embryos ($n > 200$) as described (Appelbaum et al., 2004).

Fish maintenance and examination of embryos: adult zebrafish were raised under standard conditions (Westerfield, 2000). To produce embryos, male and female zebrafish were paired in the evening, and spawning occurred within 1 h of lights on. Injected embryos were raised in a light-controlled incubator at 28°C. EGFP fluorescence in live embryos/larvae was monitored on days 2–5 of development under a stereo dissecting microscope (SZX12, Olympus, Melville, NY) equipped with appropriate set of filters (excitation, 460–490 nm; emission, 510–550 nm). Embryos were sorted on the basis of their expression pattern and results were subjected to χ^2 -analysis.

***In vitro* transient transfection assays of altered PCEs-containing constructs:** all three PCEs in the *AA1M-PRDM* construct were altered (TAATC to CAATC) using the QuickChange site-directed mutagenesis kit (see above). The presence of the mutations in the resulting construct, *AA1M-PRDM-AP3* was confirmed by sequencing. NIH 3T3 cells (CRL-1658, ATCC, Manassas, VA) were plated at a density of 3×10^4 cells per well in a 24-well plate (Costar, Cambridge, MA) and transfected 24 h later with a mixture containing Lipofectamine Plus (1.25/2.5) reagents (Invitrogen, San Diego, CA) and 10 ng of one of the following luciferase reporter vectors: the 'wild-type' *AA1M-PRDM* (Appelbaum et al., 2005); *AA1M-PRDM-MP3* (Appelbaum et al., 2005); or *AA1M-PRDM-AP3*. In addition, 0.75 μ g of a 1:1:1 expression vector mix (zBMAL2/zCLOCK2/zOTX5) or empty vector pcDNA (Invitrogen) were added. The entire mix was prepared in 50 μ l of Vitacell Dulbecco's modified Eagle's medium (ATCC) without fetal bovine serum. On the following day, the medium containing the complexes was replaced with fresh 0.5 ml of culture medium (Vitacell Dulbecco's modified Eagle's medium, supplemented with 10% fetal bovine serum); plates were harvested 24 h later. Differences in transfection efficiency were taken into account by measuring the enzyme activity generated by a co-transfected thymidine kinase promoter-driven Renilla luciferase plasmid (0.5 ng). Firefly and Renilla luciferase enzyme activities were measured using the Stop and Glo kit (Promega, Madison, WI) following the manufacturer's instructions, and relative luciferase activity was determined for each well. All experiments were performed in four replications and the data was subjected to the Student's *t*-test analysis.

3 RESULTS AND DISCUSSION

3.1 A list of transcripts highly expressed in the pineal

Using a genome-wide approach, we set out to identify transcripts that are highly expressed in the pineal gland. Our assumption was that such a list of transcripts will serve as a basis to identify *cis*-regulatory elements that confer expression of genes in the pineal gland.

To that end, we have collected multiple samples from adult transgenic zebrafish expressing EGFP in the pineal gland. This approach allowed selective removal of the pineal gland without contamination by brain tissue. Labeled RNAs were prepared and hybridized to Affymetrix chips providing a genome-wide examination of the expression level of ~15 000 transcripts. The pineal gland microarray data was compared with that of the brain microarrays (without the pineal gland and eyes) and Affymetrix microarray data of zebrafish muscle, heart and embryonic stages (prior to pineal formation) that was downloaded from NCBI GEO database (see Section 2).

The computational detection of transcripts that are highly expressed in the pineal gland was done in the following way.

- (1) The entire microarray dataset, logarithmically transformed, was normalized using quantile normalization in order to

guarantee that the distribution of probe intensities is the same in all the chips (Bolstad *et al.*, 2003).

- (2) The average expression level of each transcript in each tissue was calculated from all microarrays available for the specific tissue.
- (3) The Student's *t*-test was used in order to identify transcripts that have higher average expression level in the pineal compared with their average expression levels in each of the other tissues: brain, muscle, heart and whole embryo (four different tests for each transcript). Significance was set at Bonferroni corrected $P < 0.05$.
- (4) For each transcript that passed the mentioned criteria, the average expression level in the pineal gland was divided by the average expression levels in each of the other tissues, giving four values of fold difference for each transcript. The minimum fold difference (MFD) for each transcript was recorded.
- (5) A list which ranged transcripts by MFD was prepared. Forty transcripts with the highest MFD values are presented in the Supplementary Table.

To validate the expression pattern of the transcripts in the list (Supplementary Table), we have collected data of whole mount *in situ* hybridization (ISH) at 4–5 days post-fertilization (dpf). The spatial expression pattern of twenty of the known genes has been previously analyzed, and ISH images are available. 19 of these 20 genes exhibit high expression in the pineal (see Supplementary Table), indicating a low false-positive rate in our analysis (~5%).

3.2 Regulatory elements that drive pineal expression

With the basic assumption that a similar expression pattern reflects similar regulatory mechanisms, we sought to identify common *cis*-acting regulatory elements. The procedure used was a modification of the motif discovery workflow as suggested by MacIsaac and Fraenkel (2006).

- (1) Of the list of 40 transcripts highly expressed in the pineal, 21 promoters were identified (see Section 2 and Supplementary Table).
- (2) Three motif detection programs were used, Weeder (Pavesi *et al.*, 2004), Bioprospector (Liu *et al.*, 2001) and Oligo-analysis (van Helden *et al.*, 1998), with the list of the 21 promoters and a background model for the promoter regions as an input. The top scoring motifs with their corresponding scores were recorded for each program.
- (3) The motifs from all programs were given a *P*-value, estimated through randomization (20 000 realizations of 21 randomly chosen promoters, see Section 2). Multiple testing was accounted for by setting the false discovery rate at 20% (Benjamini and Hochberg, 1995), resulting in 17 detected motifs.
- (4) To find strong integrated patterns, the motifs from the three *ab initio* motif detection programs were clustered together using pattern-assembly tool (<http://rsat.scmbb.ulb.ac.be/rsat/>).

This procedure resulted with the identification of two motif-clusters: TAATC (five motifs) and CAATC (six motifs). TAATC was also

Table 1. Transient expression pattern of the *aanat2* promoter-EGFP construct, with and without mutation in the CAATC motif

	Wild type	Mutated
Number of zygotes injected	202	340
Number of larvae showing expression in pineal (%)	79 (39)	71 (21)
Number of larvae not showing expression in pineal (%)	123 (61)	269 (79)

identified by the program TRES as the highest ranking motif, strengthening the findings obtained by the *ab initio* motif detection programs. TRES finds enriched known binding sites in a given set of promoters (Katti *et al.*, 2000).

The abundance of these motifs in the given list of promoters is about 2-fold higher than expected for a randomly chosen list of zebrafish promoters: TAATC and CAATC have 58 and 41 occurrences compared with the expected number of 30 and 21, respectively ($P < 0.05$, Poisson distribution, Bonferroni corrected). The two putative motifs are well distributed in the list of promoters: TAATC appears (at least once) in all the promoters tested and CAATC appears in 20 out of 21 promoters.

The same motif detection procedure was performed on an extended list of transcripts with high-pineal expression (MFD > 5, ~80 genes). CAATC and TAATC were again detected using the extended list (along with three new motifs that were not further analyzed), indicating that the motif detection procedure is robust.

The best established DNA binding site known to drive gene expression in the pineal gland is the PCE—TAATC/T (Appelbaum and Gothilf, 2006). In addition to re-discovering the PCE, our approach also identified CAATC, which has not been previously reported as being associated with expression of genes in photoreceptor tissues.

3.3 CAATC has a role in pineal expression *in vivo*

The role of CAATC was experimentally evaluated *in vivo* by point mutations of promoter-EGFP constructs followed by microinjection into fertilized eggs and transient expression in developing embryos (Supplementary Fig. 1). The promoter used is that of the pineal-specific *aanat2* (Appelbaum *et al.*, 2004) which includes the sequence CAATC in location -320 bp with respect to the TSS. This sequence was mutated (CAATC to CGGTC), the resulting construct was microinjected to one-cell stage embryos, and the expression pattern was compared with that of the original construct which was used as a control (Section 2). The mutated construct showed a significant (~50%, $P < 0.001$, χ^2) reduction in pineal-gland expression of EGFP as compared with that of the normal construct (see Table 1 and Section 2). These results indicate that CAATC is important in driving expression of *aanat2* in the pineal gland.

3.4 CAATC and PCE are functionally different

In light of the clear sequence resemblance of the CAATC motif to PCE (TAATC) and the fact that both elements facilitate gene expression in the pineal gland, it was reasonable to suspect that the CAATC motif is an extension of the PCE element.

Had this been the case, one would have expected OTX5 to bind CAATC as well as TAATC. In order to check this possibility, *in vitro* transfection analyses of modified AA1M-PRDM constructs in combinations with OTX5 and BMAL/CLOCK expression vectors were performed. The AA1M-PRDM construct is a luciferase vector driven by a minimal arginine vasopressin (AVP) promoter and the PRDM enhancer. This enhancer was previously found to be important for pineal-specific gene expression due to the fact that it contains three functional copies of the PCE and a functional E-box (Appelbaum *et al.*, 2004). OTX5 and BMAL/CLOCK were shown to synergistically induce expression via these elements (Appelbaum *et al.*, 2005). Three versions of the PRDM construct were used in the current experiment: (i) wild type; (ii) a mutated version, AA1M-PRDM-MP3, in which all three PCEs were mutated and are not functional (Appelbaum *et al.*, 2005); and (iii) an altered version, AA1M-PRDM-AP3, in which all three PCE were changed to CAATC (Section 2).

As expected, transfection of OTX5 significantly ($P < 1e-3$, *t*-test) increased the expression driven by the AA1M-PRDM construct in comparison with the mutated construct AA1M-PRDM-MP3. Importantly, AA1M-PRDM-AP3 gives slightly lower expression level than AA1M-PRDM-MP3 when co-transfected with OTX5, indicating that the CAATC element does not mediate OTX5 actions (Supplementary Fig. 2).

Co-transfection of OTX5 with BMAL/CLOCK further increased (~7-fold) the expression driven by the AA1M-PRDM construct, reflecting the synergistic nature of OTX5 and BMAL/CLOCK (Appelbaum *et al.*, 2005). This increase in activity was significantly higher ($P < 1e-4$, *t*-test) than the induction of AA1M-PRDM-MP3 and AA1M-PRDM-AP3 by the combination of OTX5 and BMAL/CLOCK (Supplementary Fig. 2). These results further indicate that the CAATC element does not mediate the action of OTX5, demonstrating that the novel CAATC element is functionally different from the PCE and probably serves as a DNA binding site for a different transcription factor.

Interestingly, the CAATC motif is a string within PIPE which is known to mediate OTX5-independent pineal-enhanced transcription (Asaoka *et al.*, 2002). However, the consensus of all the occurrences of CAATC motif in the promoters of genes that are highly expressed in the zebrafish pineal gland—WNNCAATCNNWYNW—shows no other sequence resemblance to PIPE, thereby suggesting that these two motifs are different. Further experimental work will determine the relationships of these two motifs.

Although this work provides strong evidence that CAATC is important in conferring expression of genes in the pineal gland of the zebrafish, it should be added that we have not been able to identify CAATC as a conserved feature in the promoters of genes that are highly expressed in the pineal gland of rat (S.Alon, unpublished). Accordingly, it is possible that CAATC may reflect an evolutionary development that occurred following the emergence of fish from the line leading to mammals.

Together, this study implicates the involvement of a novel motif (CAATC) in determining enhanced expression of genes in the zebrafish pineal gland. This information may be utilized to identify yet another transcription factor that contributes to pineal expression and therefore to a better understanding of pineal determination.

Given the reliability of the methodology described here for the identification of *cis*-acting regulatory elements in zebrafish and the relative ease of performing *in vivo* mutation analysis in this species,

this approach may be applicable for finding additional regulatory elements governing specific gene expression in other tissues.

Funding: United States-Israel Binational Science Foundation, Jerusalem (2005280).

Conflict of Interest: none declared.

REFERENCES

- Appelbaum, L. *et al.* (2004) Zebrafish serotonin-N-acetyltransferase-2 gene regulation: pineal-restrictive downstream module contains a functional E-box and three photoreceptor conserved elements. *Mol. Endocrinol.*, **18**, 1210–1221.
- Appelbaum, L. *et al.* (2005) Homeobox-clock protein interaction in zebrafish. A shared mechanism for pineal-specific and circadian gene expression. *J. Biol. Chem.*, **280**, 11544–11551.
- Appelbaum, L. and Gothilf, Y. (2006) Mechanism of pineal-specific gene expression: the role of E-box and photoreceptor conserved elements. *Mol. Cell. Endocrinol.*, **252**, 27–33.
- Asaoka, Y. *et al.* (2002) Pineal expression-promoting element (PIPE), a *cis*-acting element, directs pineal-specific gene expression in zebrafish. *Proc. Natl Acad. Sci. USA*, **99**, 15456–15461.
- Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B*, **57**, 289–300.
- Blanchette, M. *et al.* (2006) Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression. *Genome Res.*, **16**, 656–668.
- Bolstad, B.M. *et al.* (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.
- Chen, H. *et al.* (2007) Expression of the G protein γ T1 subunit during zebrafish development. *Gene Expr. Patterns*, **7**, 574–583.
- Davies, S.R. *et al.* (2007) Computational identification and functional validation of regulatory motifs in cartilage-expressed genes. *Genome Res.*, **17**, 1438–1447.
- Falcon, J. (1999) Cellular circadian clocks in the pineal. *Prog. Neurobiol.*, **58**, 121–162.
- Gamse, J.T. *et al.* (2002) Otx5 regulates genes that show circadian expression in the zebrafish pineal complex. *Nat. Genet.*, **30**, 117–121.
- Gothilf, Y. *et al.* (2002) Pineal-specific expression of green fluorescent protein under the control of the serotonin-N-acetyltransferase gene regulatory regions in transgenic zebrafish. *Dev. Dyn.*, **225**, 241–249.
- GuhaThakurta, D. (2006) Computational identification of transcriptional regulatory elements in DNA sequence. *Nucleic Acids Res.*, **34**, 3585–3598.
- GuhaThakurta, D. *et al.* (2002) Identification of a novel *cis*-regulatory element involved in the heat shock response in *Caenorhabditis elegans* using microarray gene expression and computational methods. *Genome Res.*, **12**, 701–712.
- Katti, M.V. *et al.* (2000) TRES: comparative promoter sequence analysis. *Bioinformatics*, **16**, 739–740.
- Li, X. *et al.* (1998) A pineal regulatory element (PIRE) mediates transactivation by the pineal/retina-specific transcription factor CRX. *Proc. Natl Acad. Sci. USA*, **95**, 1876–1881.
- Liu, X. *et al.* (2001) BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac. Symp. Biocomput.*, **6**, 127–138.
- MacIsaac, K.D. and Fraenkel, E. (2006) Practical strategies for discovering regulatory DNA sequence motifs. *PLoS Comput. Biol.*, **2**, e36.
- Nishio, S.I. *et al.* (2008) OTX5 regulates pineal expression of the zebrafish REV-ERB $\{\alpha\}$ through a new DNA binding site. *Mol. Endocrinol.*, **22**, 23–32.
- Pavesi, G. *et al.* (2004) Weeder web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res.*, **32**, W199–W203.
- Roth, F.P. *et al.* (1998) Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat. Biotechnol.*, **16**, 939–945.
- van Helden, J. *et al.* (1998) Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J. Mol. Biol.*, **281**, 827–842.
- Wang, H. *et al.* (2006) Experimental validation of predicted mammalian erythroid *cis*-regulatory modules. *Genome Res.*, **16**, 1480–1492.
- Westerfield, M. (2000) *The Zebrafish Book: A Guide for the Laboratory use of Zebrafish (Danio rerio)*. 4th edn. University of Oregon Press, Eugene (Oregon).